Modeling social networks

R. Lambiotte
M. Ausloos

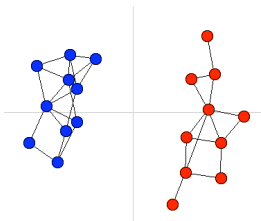
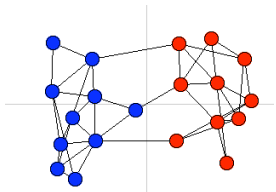
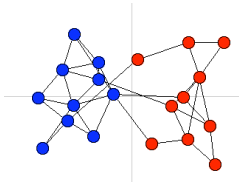
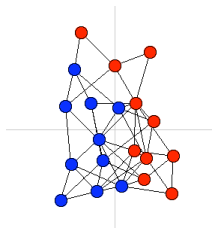
Content

I. Some background
   a) What about networks?
   b) CREEN

II. Growing networks
   a) Barabasi-Albert networks
   b) Copying each others' environment
   c) Some analytical predictions

III. Visualizing network dynamics

IV. Conclusion

Examples of complex networks:

- Internet
- Transport networks
- Protein interaction networks
- Food webs
- Social networks, e.g. collaboration networks, citation networks, friendship networks...

Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks: Peter S. Bearman, James Moody and Katherine Stovel

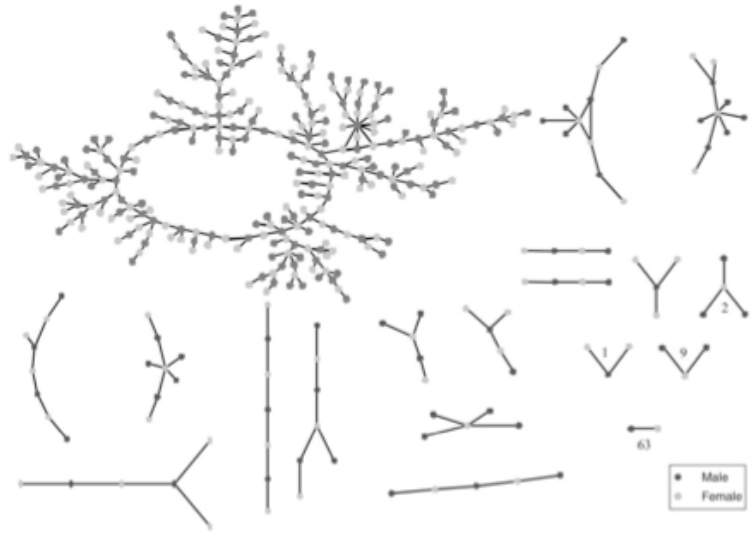
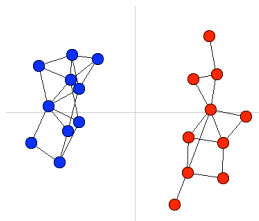
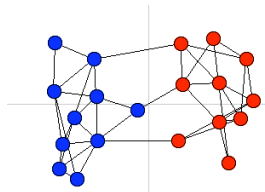
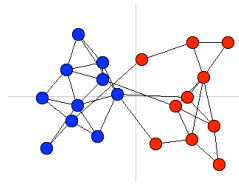
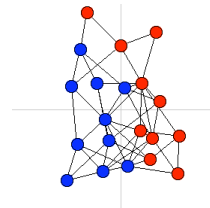


FIG. 2.—The direct relationship structure at Jefferson High



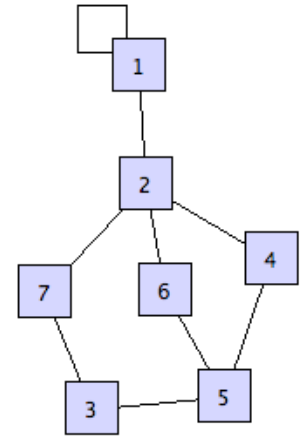




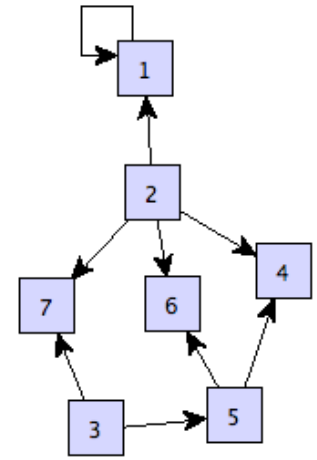


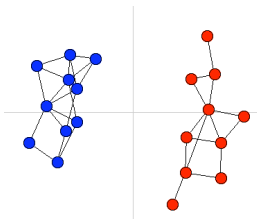
Local properties

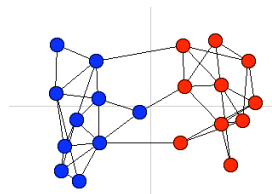
node degree: numbers of links arriving at one node

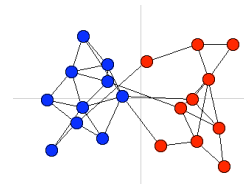


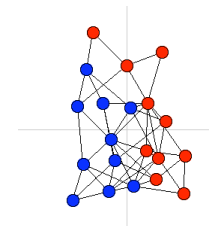
in(out) degree: number of ingoing (outgoing) links for one node (directed networks)





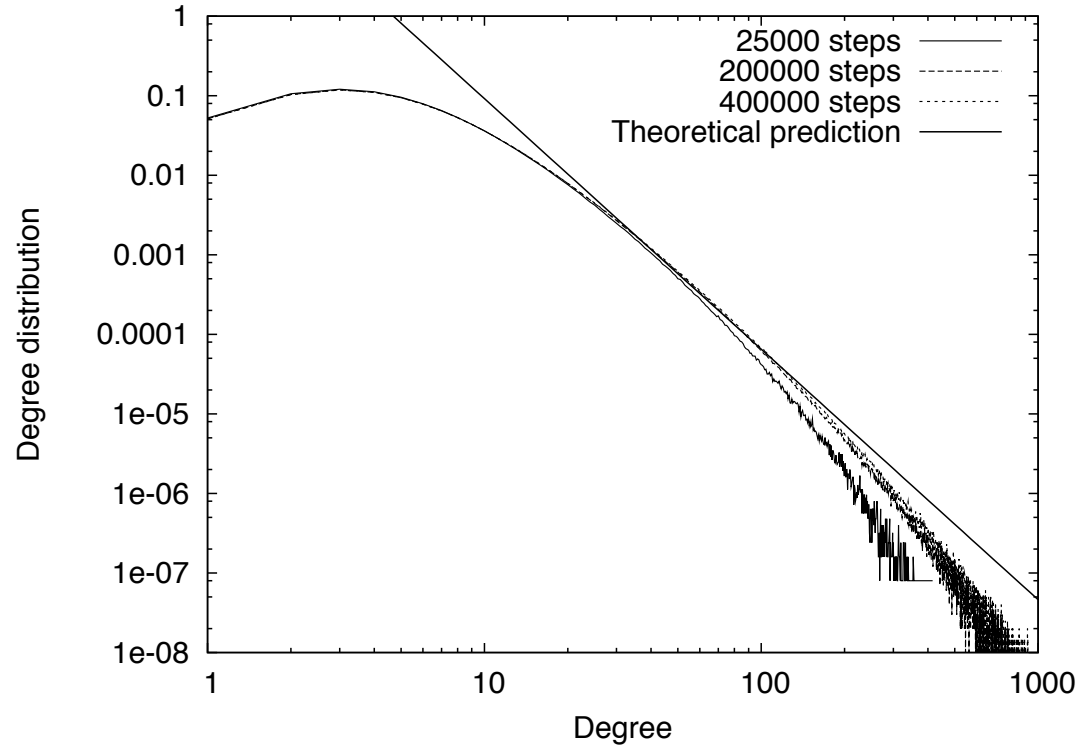






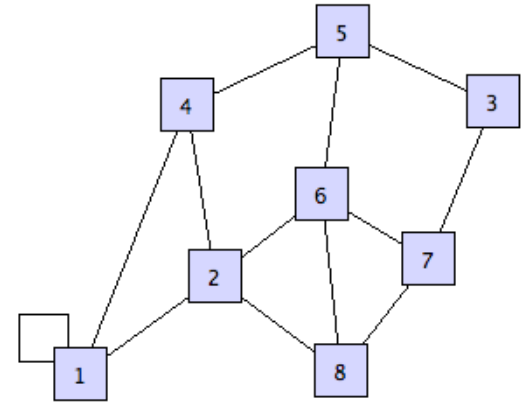
Degree distribution

the degree distribution is known to behave like power-laws in many situations, usually with an exponential cut-off



Local properties

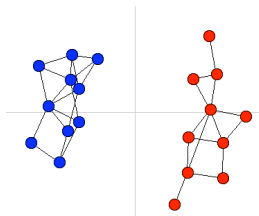
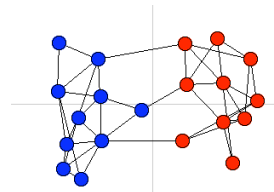
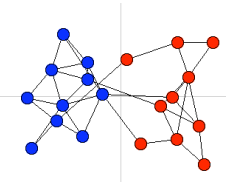
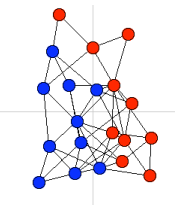
clustering coefficient: number of triangles



local correlations: assortativity...

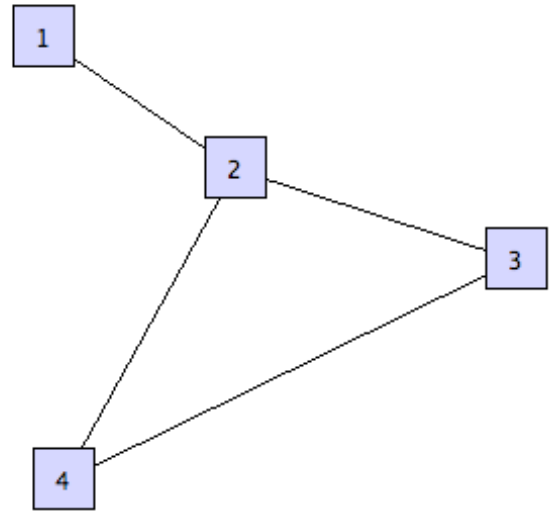
High-degree vertices prefer to attach other high-degree vertices (social networks)
High-degree vertices prefer to attach low-degree vertices (technological networks)

	network	n	r
real-world networks	physics coauthorship[a]	52 909	0.363
	biology coauthorship[a]	1 520 251	0.127
	mathematics coauthorship[b]	253 339	0.120
	film actor collaborations[c]	449 913	0.208
	company directors[d]	7 673	0.276
	Internet[e]	10 697	−0.189
	World-Wide Web[f]	269 504	−0.065
	protein interactions[g]	2 115	−0.156
	neural network[h]	307	−0.163
	food web[i]	92	−0.276
models	random graph[u]		0
	Callaway *et al.*[v]		$\delta/(1+2\delta)$
	Barabási and Albert[w]		0

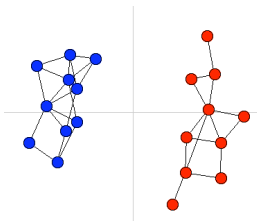
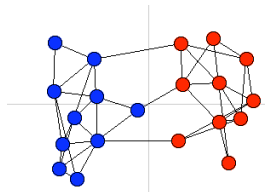
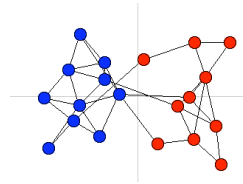
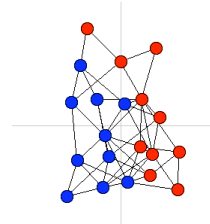





Global properties

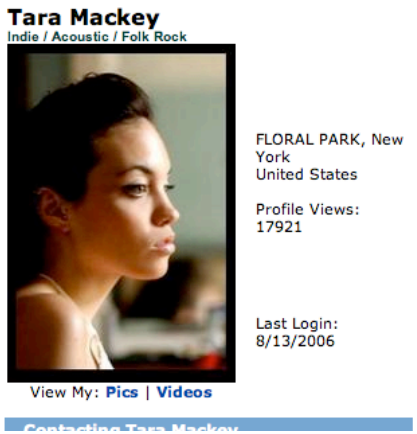
distance: length of the shortest path between nodes
small-world behaviour

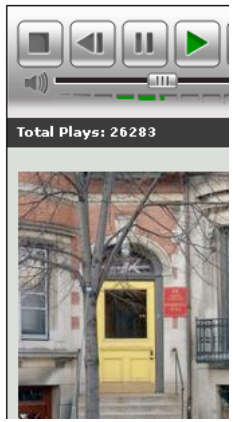


$$d \sim \ln N$$

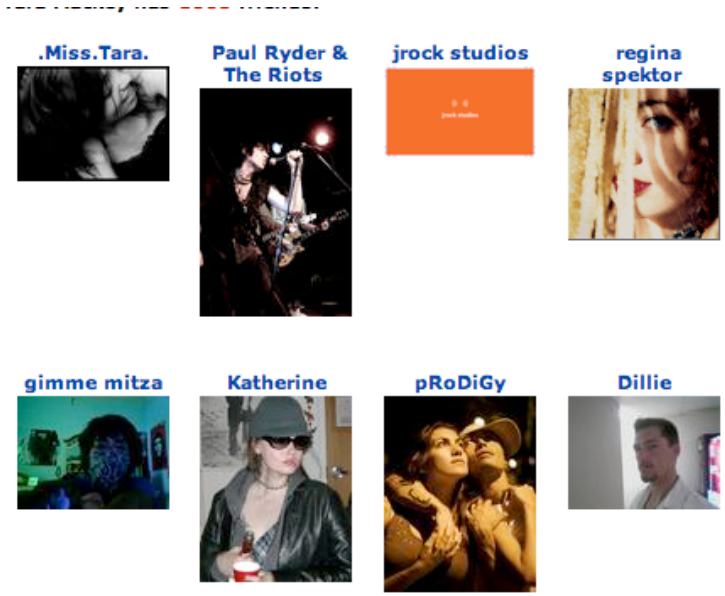





Hype of social networking





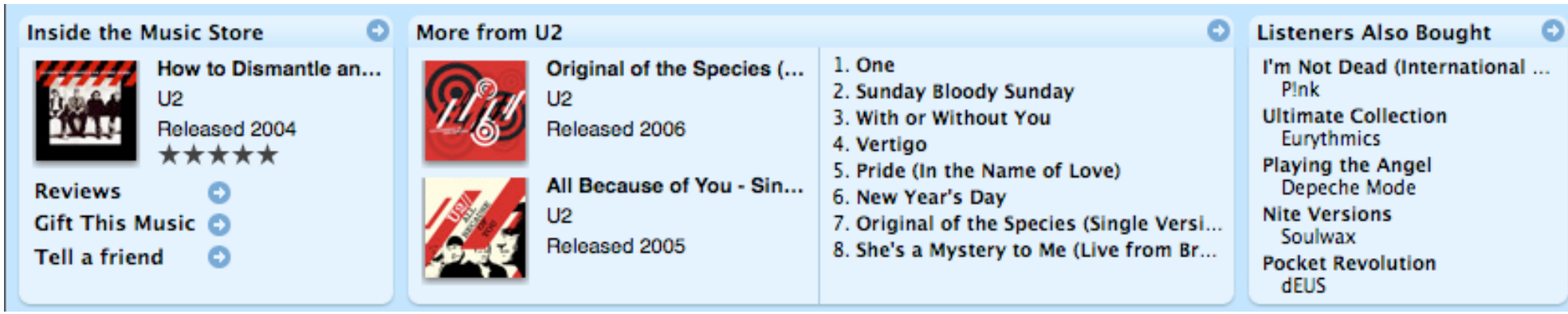
flicker, myspace, audioscrobbler, ...



New marketing/sales strategies

Importance of the buzz of a product: discussions between bloggers, chatters...

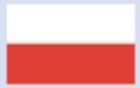
The huge amount of online databases helps to “measure” and predict the behaviour of users



CREEN

Critical Events in Evolving Networks
(www.creen.org)

Participant universities



Faculty of Physics and the Center for Complex Systems at Warsaw University of technology

SUPRATECS at University of Liège

Statistical Cybermetrics Research Group at University of Wolverhampton

Netherlands Institute for Scientific Information Services (NIWI) of the Royal Netherlands Academy of Arts and Sciences (KNAW)

Faculty of Informatics at University of Karlsruhe

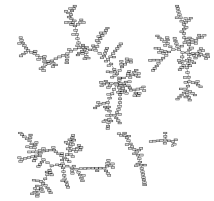
17/01/2006
Fondation Universitaire,
Rue d'Egmont, 4
1050 Ixelles

CREEN

Critical Events in Evolving Networks

This workshop aims at an inter-disciplinary audience, interested in the various ways to apprehend social networks: gathering and treatment of empirical data (especially from blogs and RSS feeds), visualization of network structures, conceptualization and network modeling by using statistical physics methods.

Program

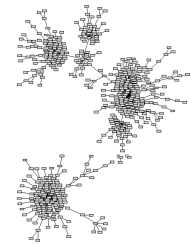


09.15-09.45: Presentation

09.45-10.30: J. Holyst, *Competing social groups in social networks*

10.30-11.15: R. Lambiotte, *Modeling growing social networks*

11.15-12.00: V. Eguiluz, *Epidemic threshold in structured scale-free networks*

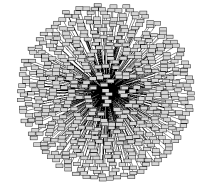


13.30-14.15: M. Thelwall, *Impact factors of scientist's personal pages*

14.15-15.00: R. Rousseau, *Lorenz curves determine partial orders for inequality and similarity measurements*

15.00-15.45: D. Wagner: Visone: *Visualizing dynamic networks*

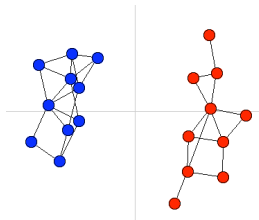
15:45-16.30: U. Brandes: *Network analysis: Methodological foundations*

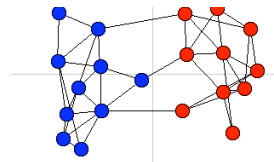


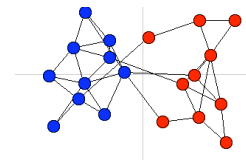
16.45-17.30: J. van Ginneken, *Hidden persuaders*

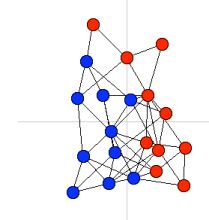
17.30-18.15: I. Hellsten & A. Scharnhorst, *Self-citations, co-authorships and keywords: : A new approach to scientists' field mobility?*

contact: R. Lambiotte, B5, Université de Liège, B-4000 LIEGE, Belgique
E-mail: renaud.lambiotte@ulg.ac.be Tel: +32 4 366 3660
Info: http://www.creen.org



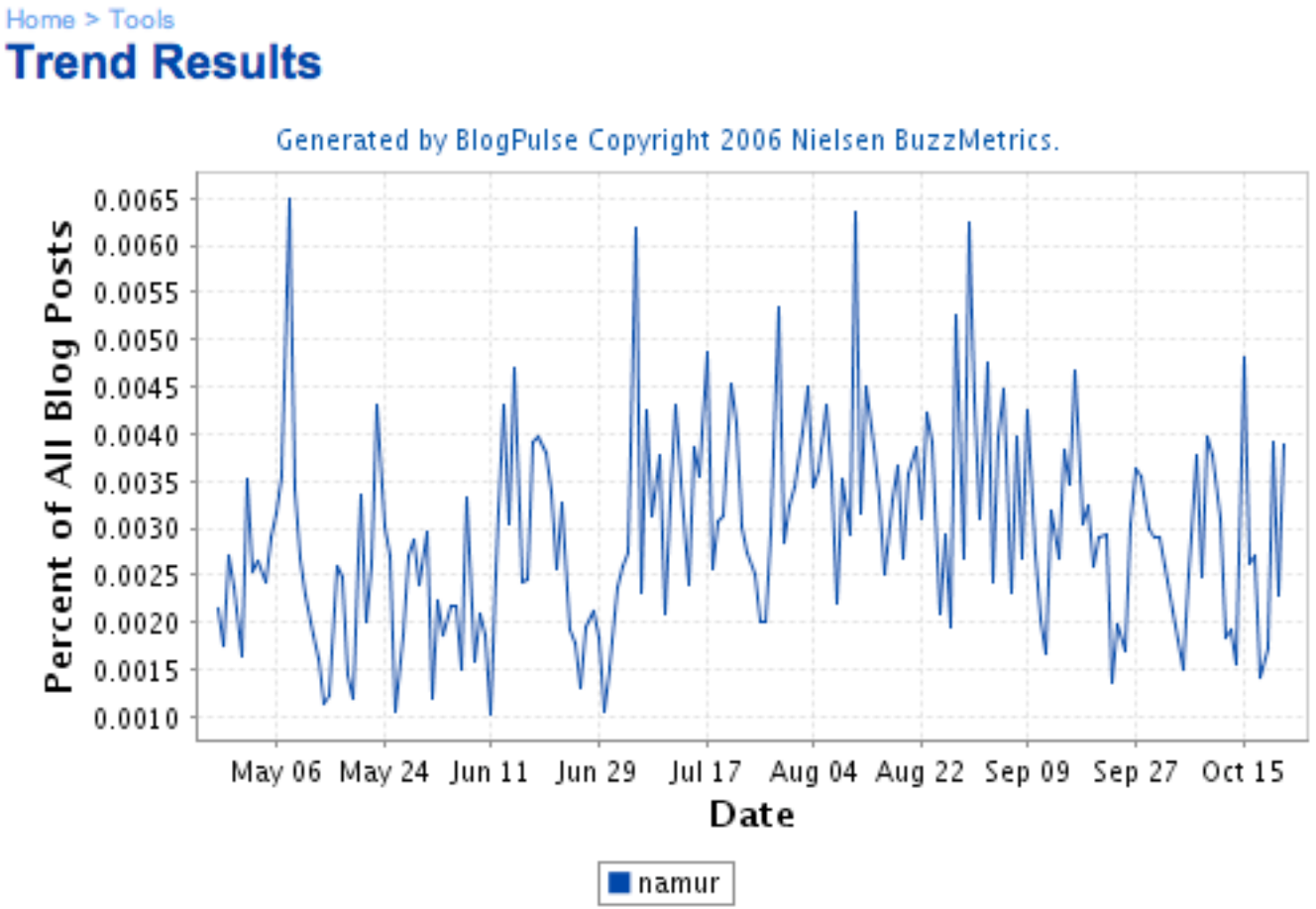






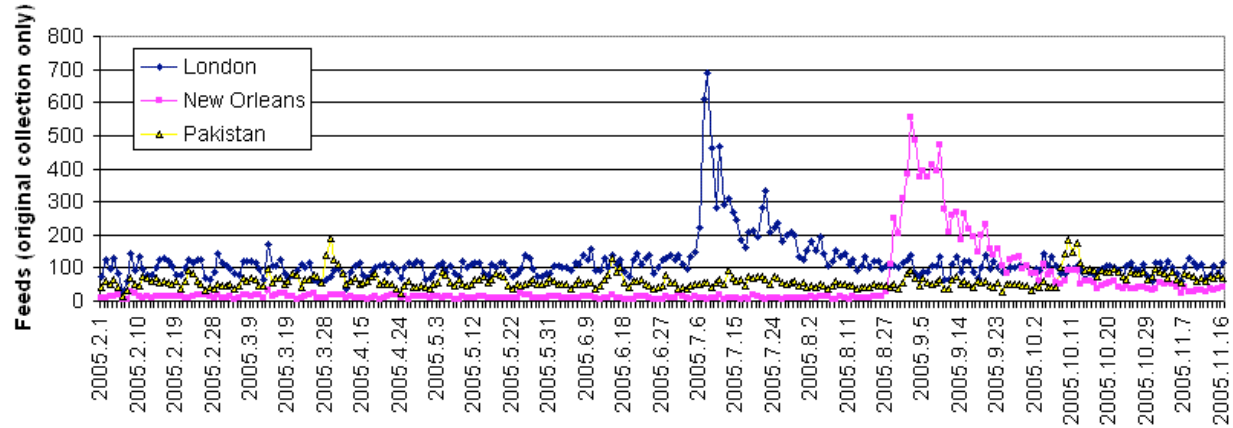
Wolverhampton

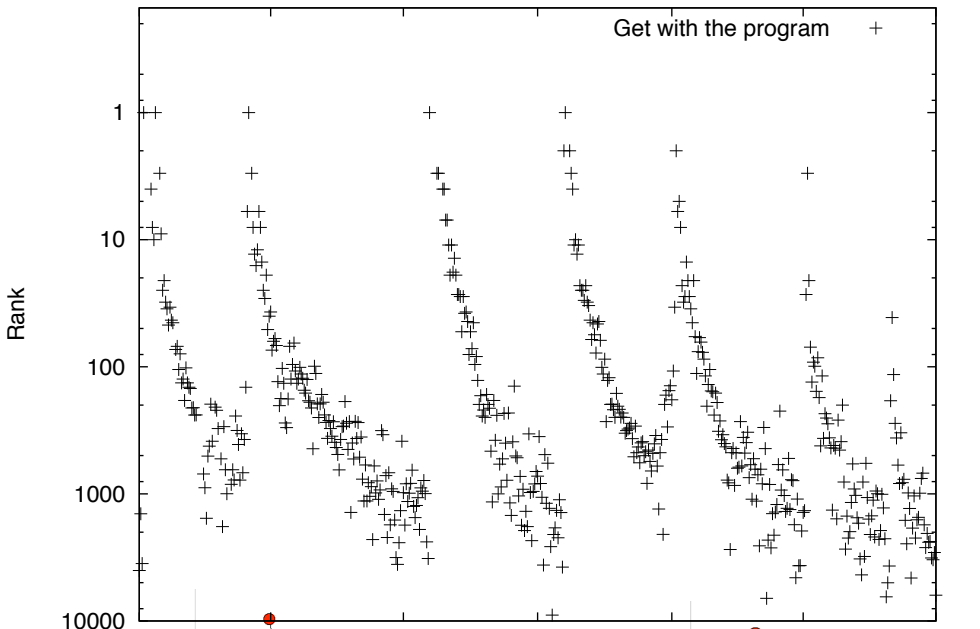
Gathering of blog data: hyperlinks, content...
(e.g. blogpulse)

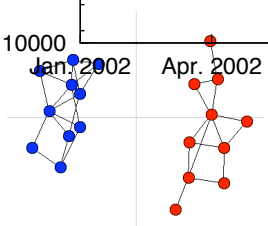


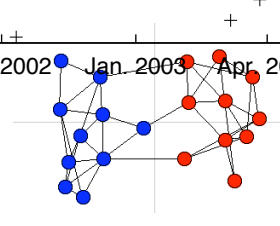
Wolverhampton

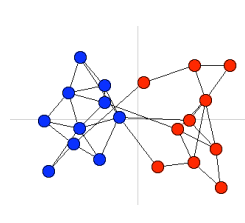
Relaxation after a shock?

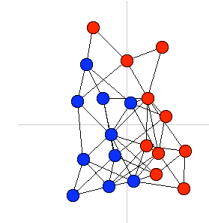






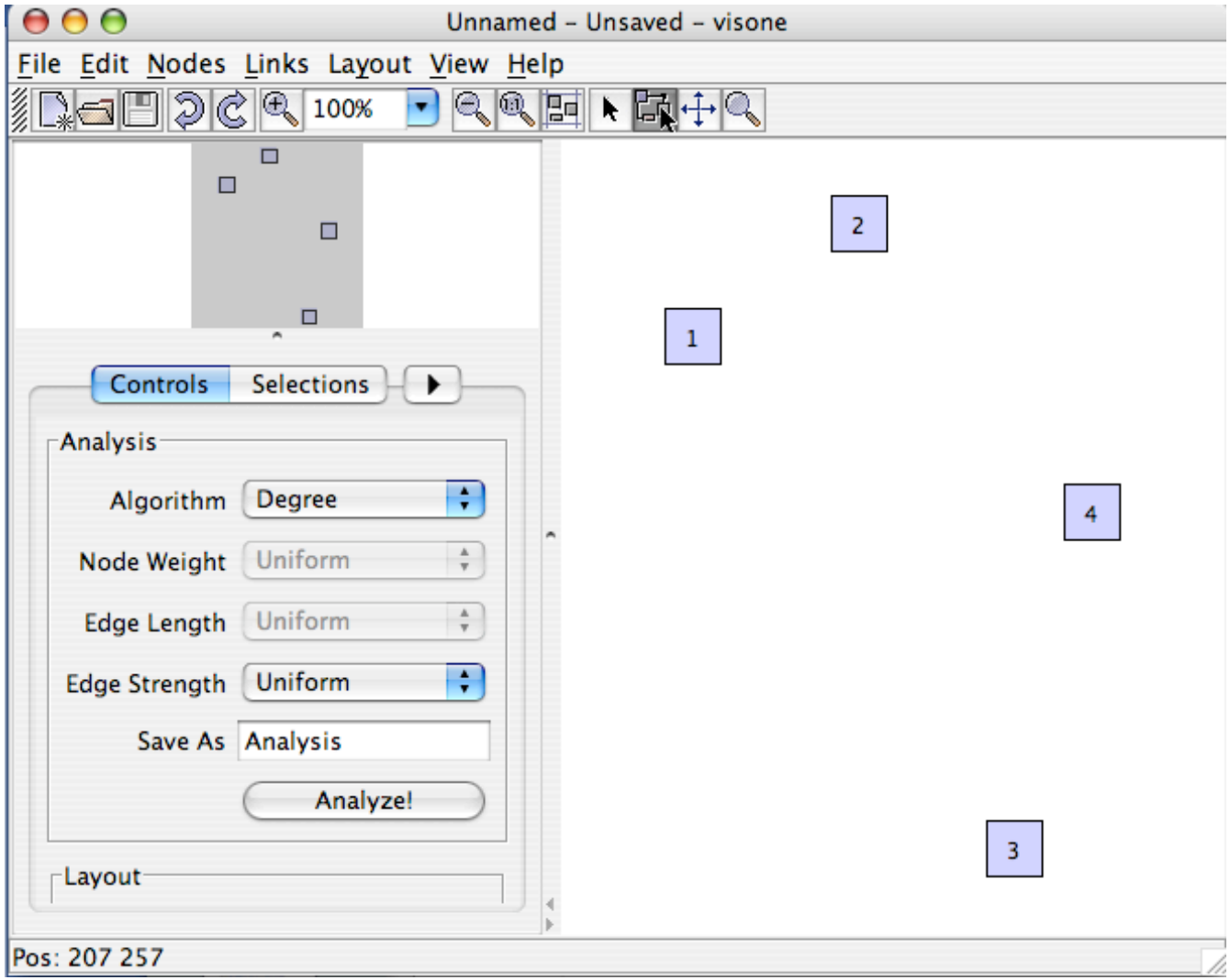






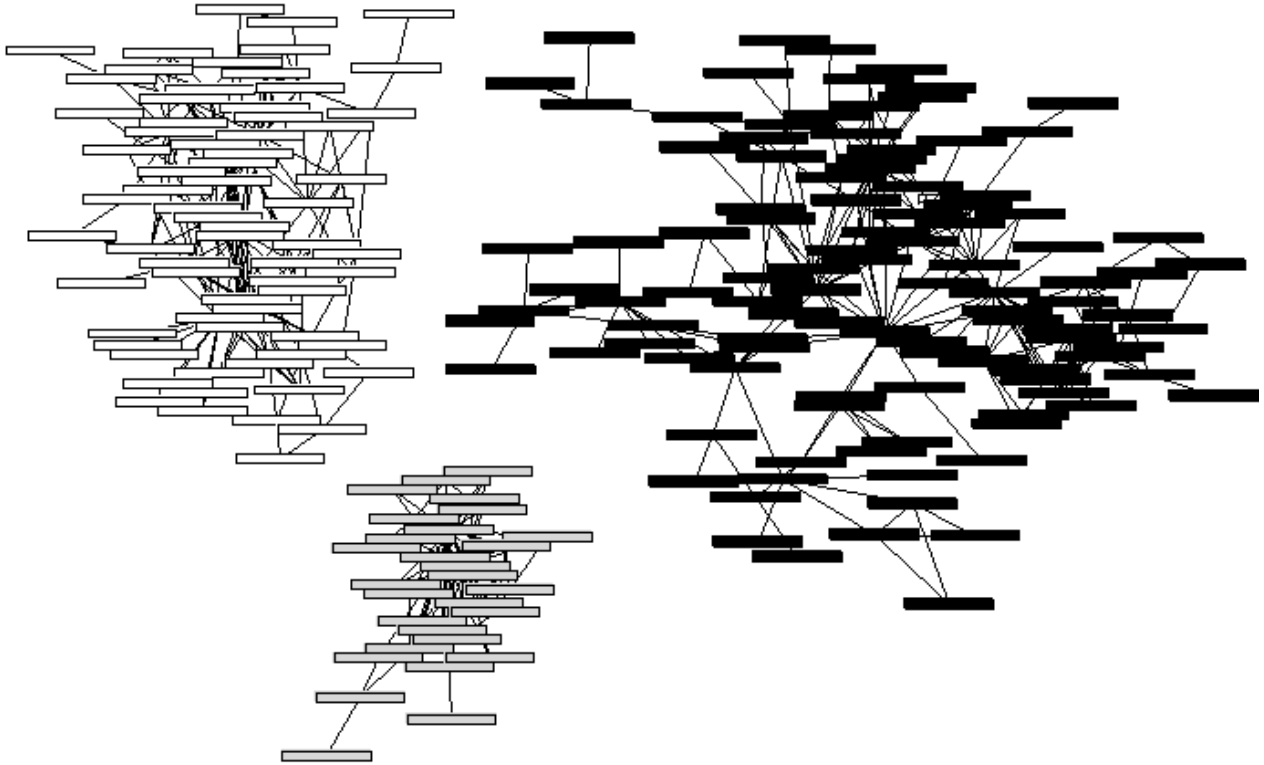
Karlsruhe

Visone: static and dynamic visualization of networks



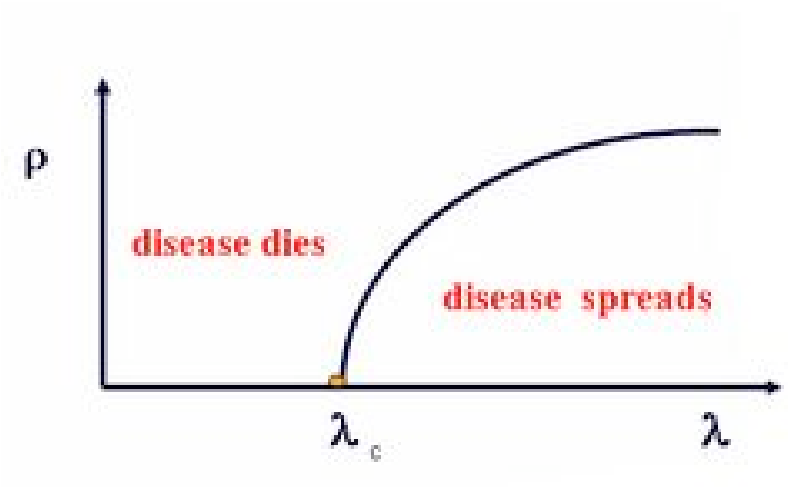
Amsterdam

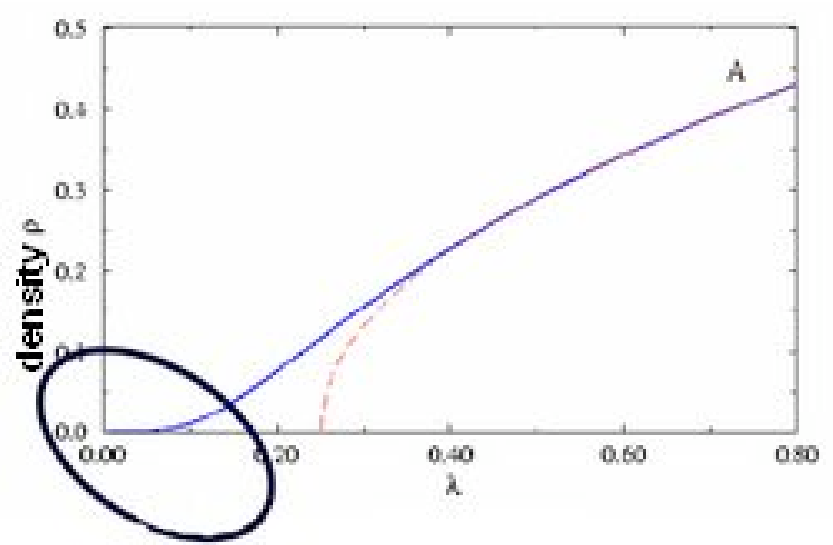
Collaboration networks: extraction of information from the network structure



Liège+Warsaw

Modeling of processes taking place on the network: virus spreading





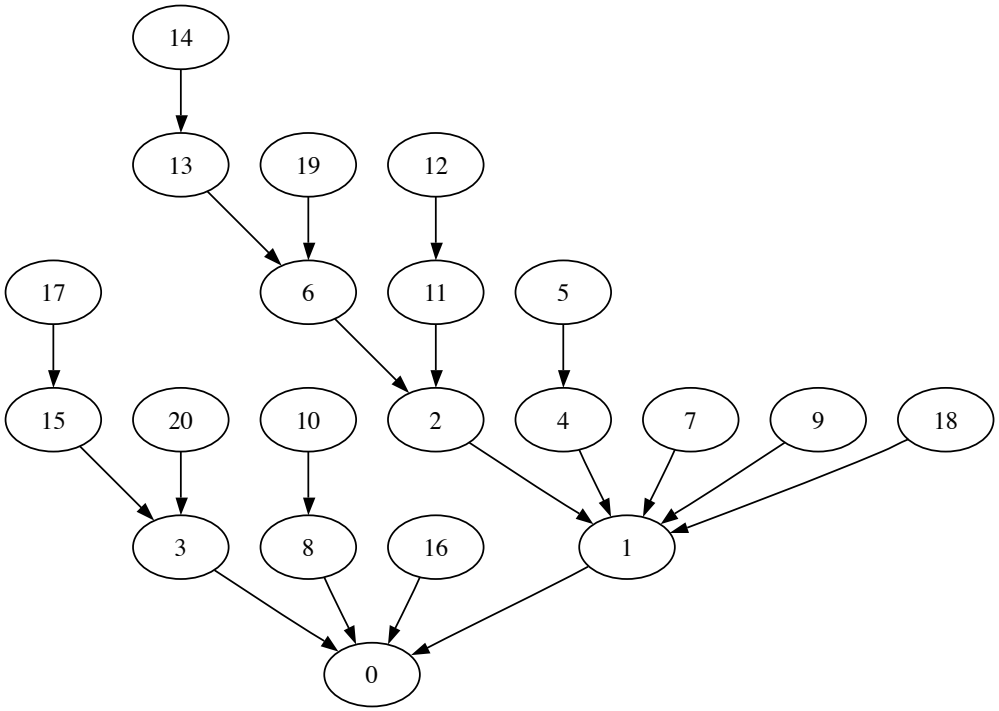
Homogeneous network

Heterogeneous network

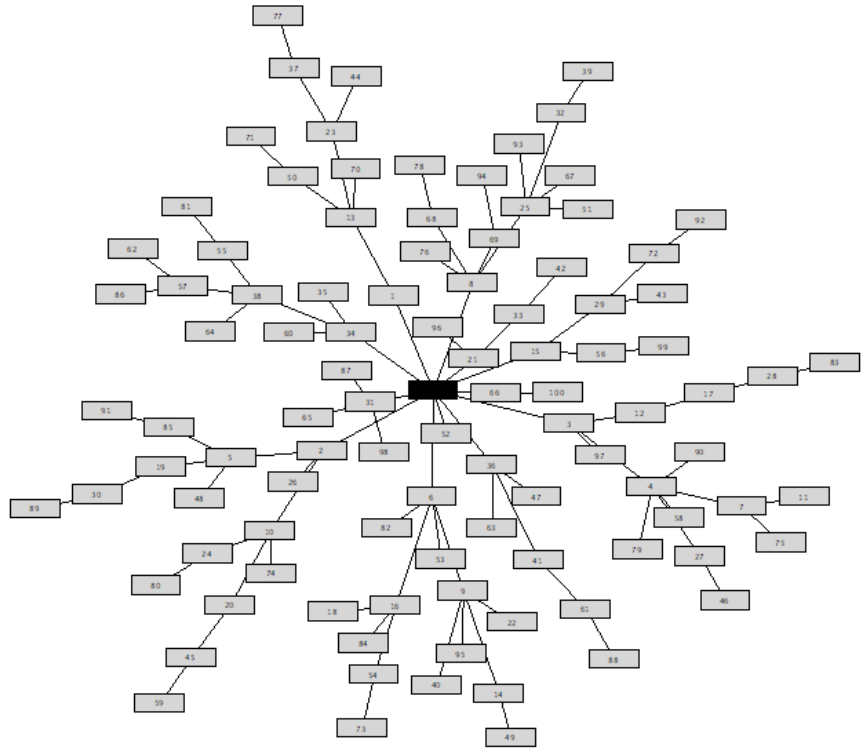
Liège+Warsaw

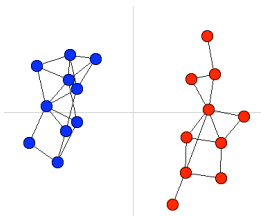
Modeling of processes how the network evolve (grow):

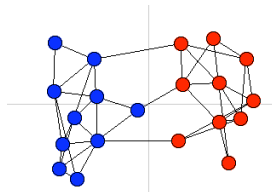
directed networks (trees)

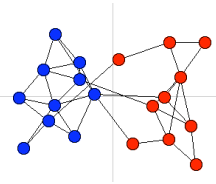


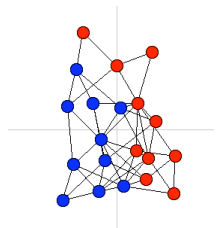
undirected networks







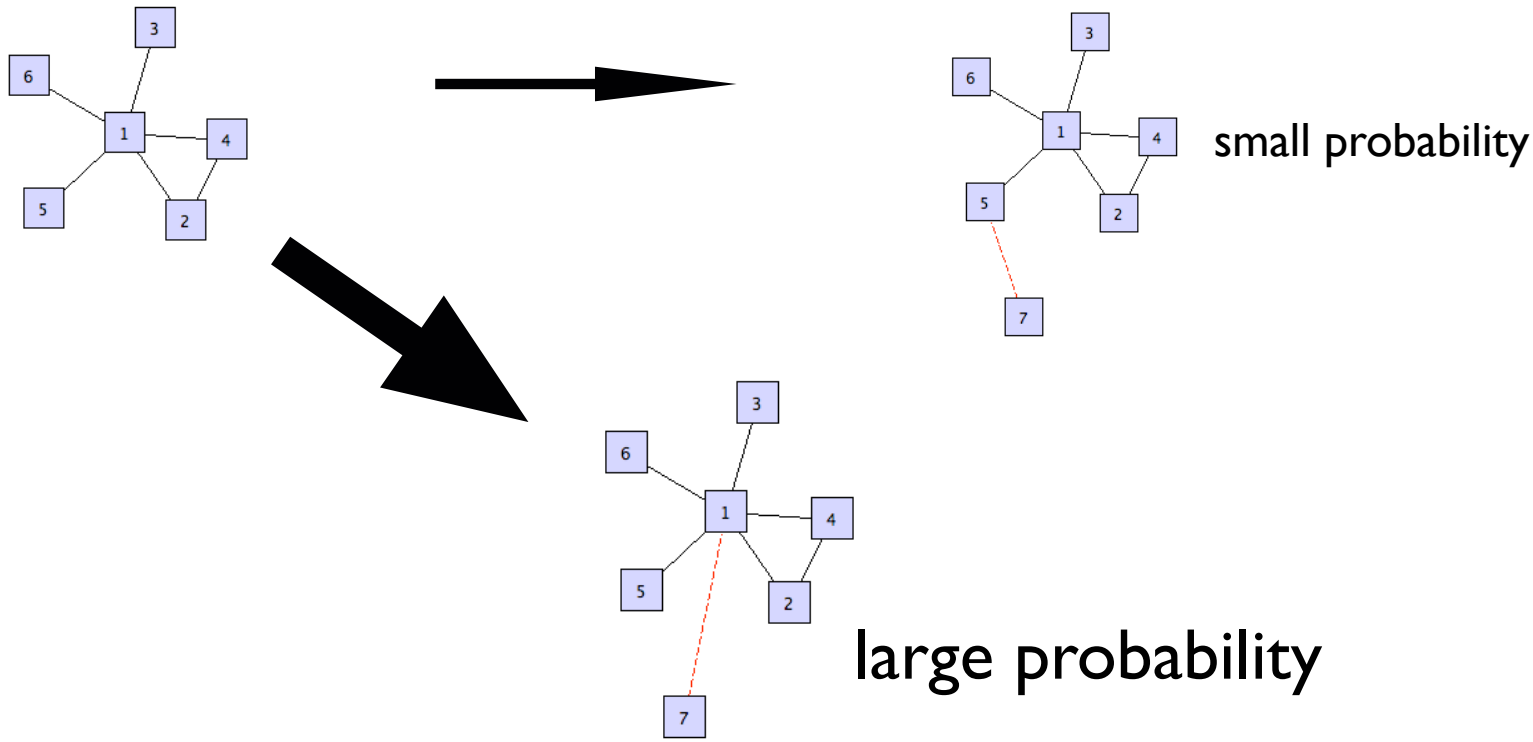




Barabasi-like networks

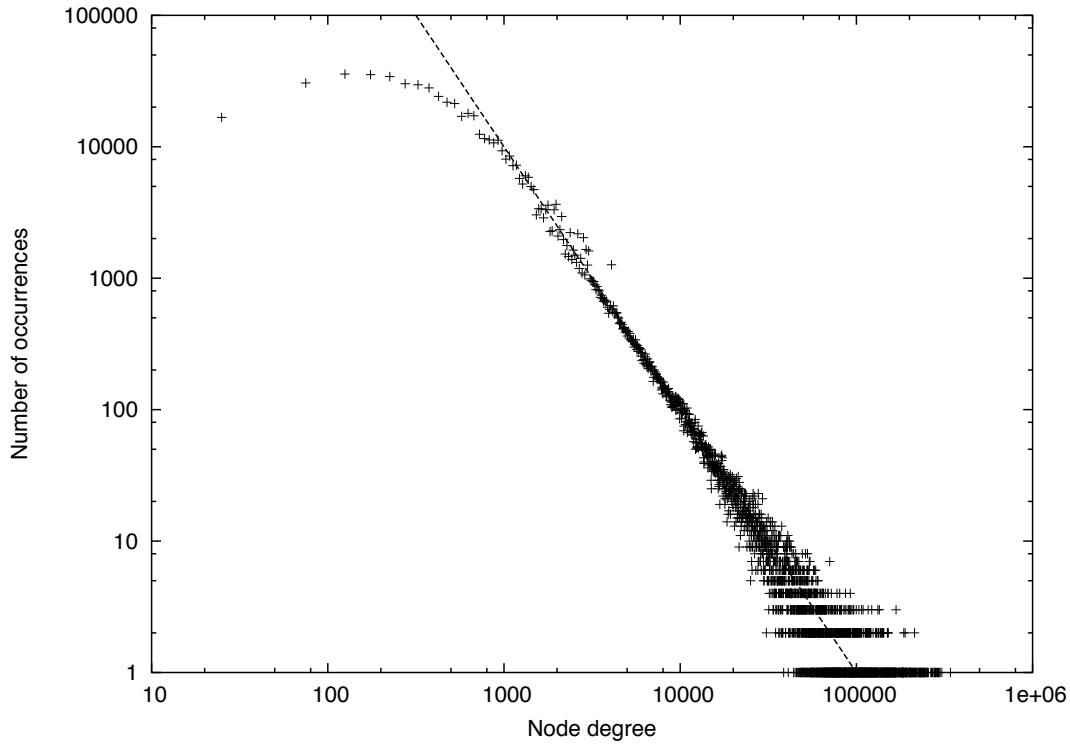
Many social systems exhibit what is called cumulated advantage, preferential attachment...

multiplicative process



This model accounts for the importance of the degree of the nodes...

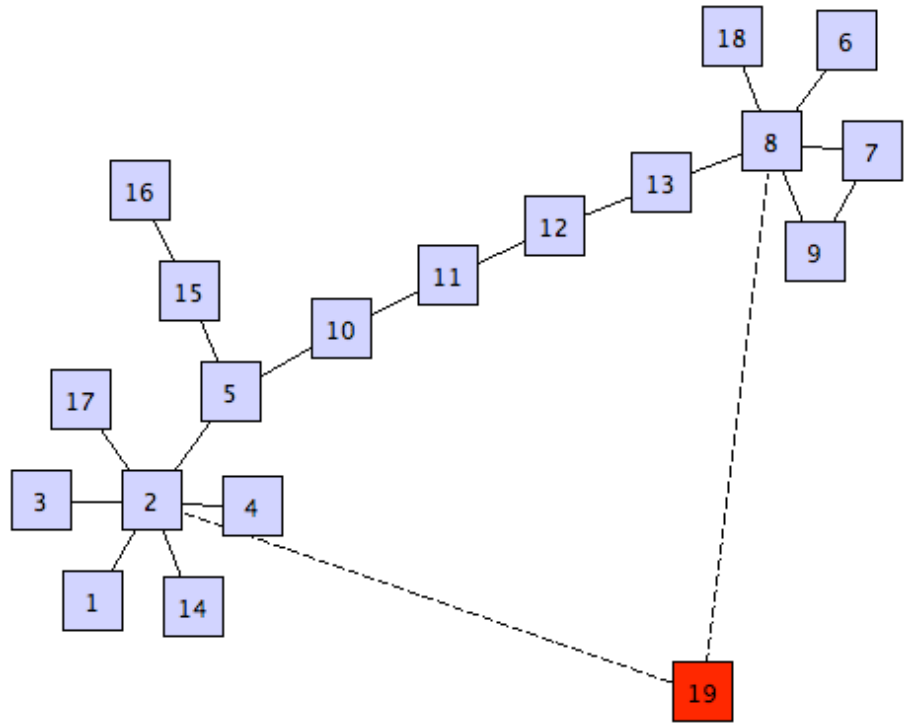
... and reproduces quite well the emergence of power-law degree distributions

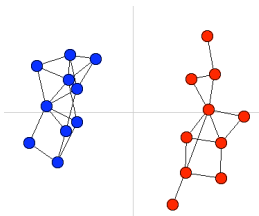


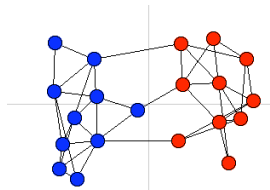
Distribution of the audience of music groups

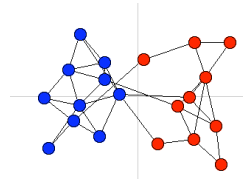
This model accounts for the importance of the degree of the nodes...

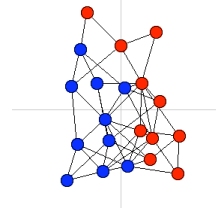
... but it is a “mean-field” model where the localization of the nodes is not taken in account (taste, kind, genre.. `





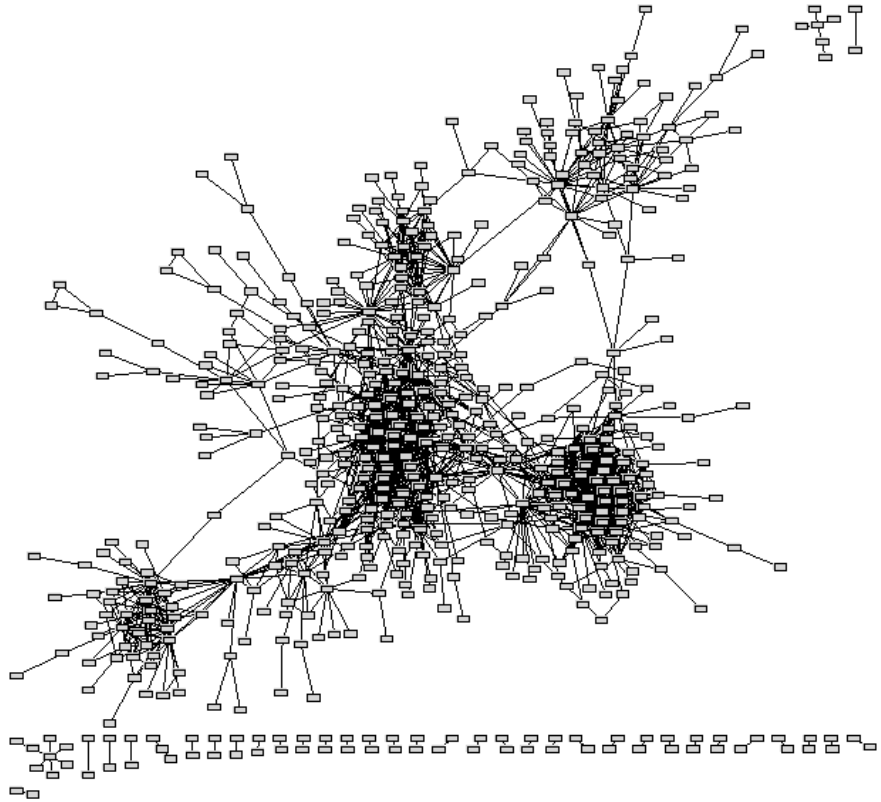




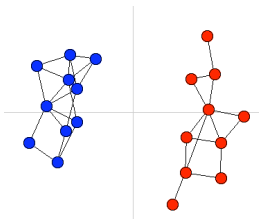


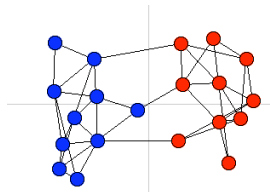
This model accounts for the importance of the degree of the nodes...

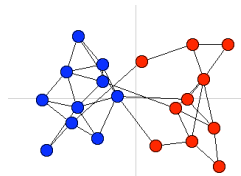
... and does not reproduces the structuring of the network (clustering coefficient, communities...)

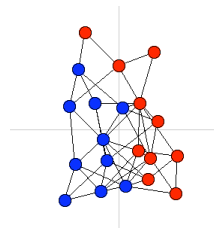


Network of music groups







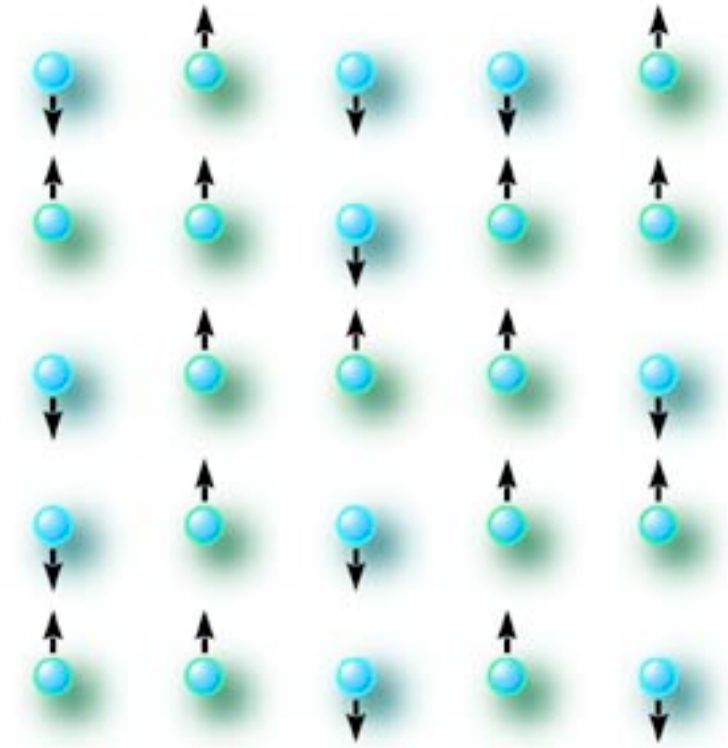


	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW `nd.edu`	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

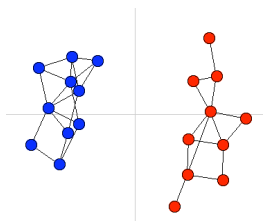
BA model: C=0, r=0, l=log(log(N))

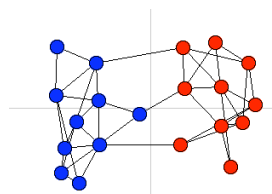
Models with local mechanisms: beyond the mean-field approach

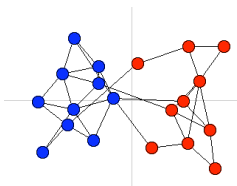
Simplest interaction possible: imitation.
You copy what you friend is doing.
E.g. listen to the same music, have the same friends, cite articles in the same scientific field...

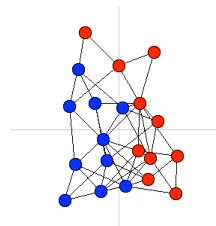


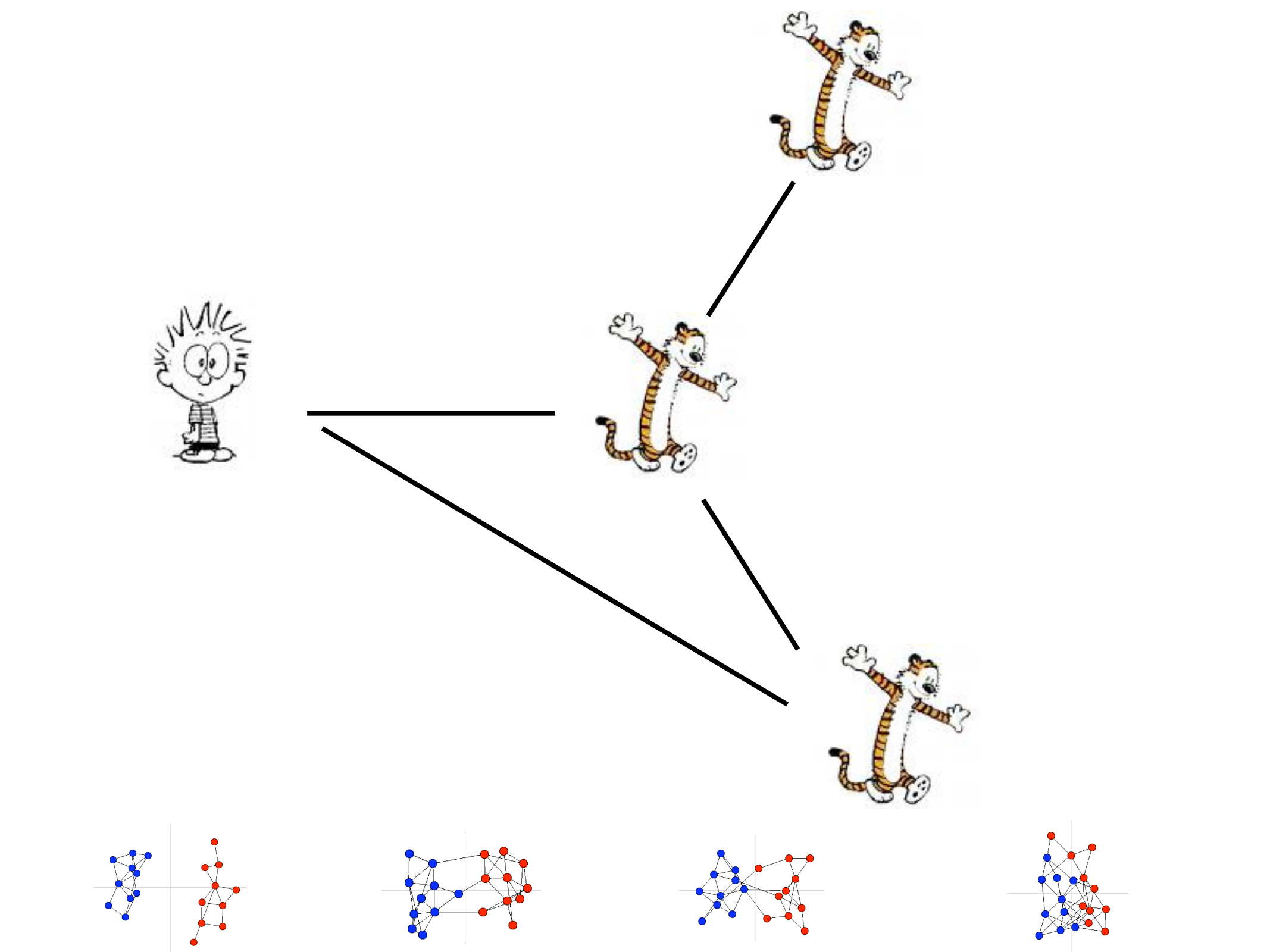
In the Ising model, spins of neighbours have a tendency to align.
Similar model for the network structure?







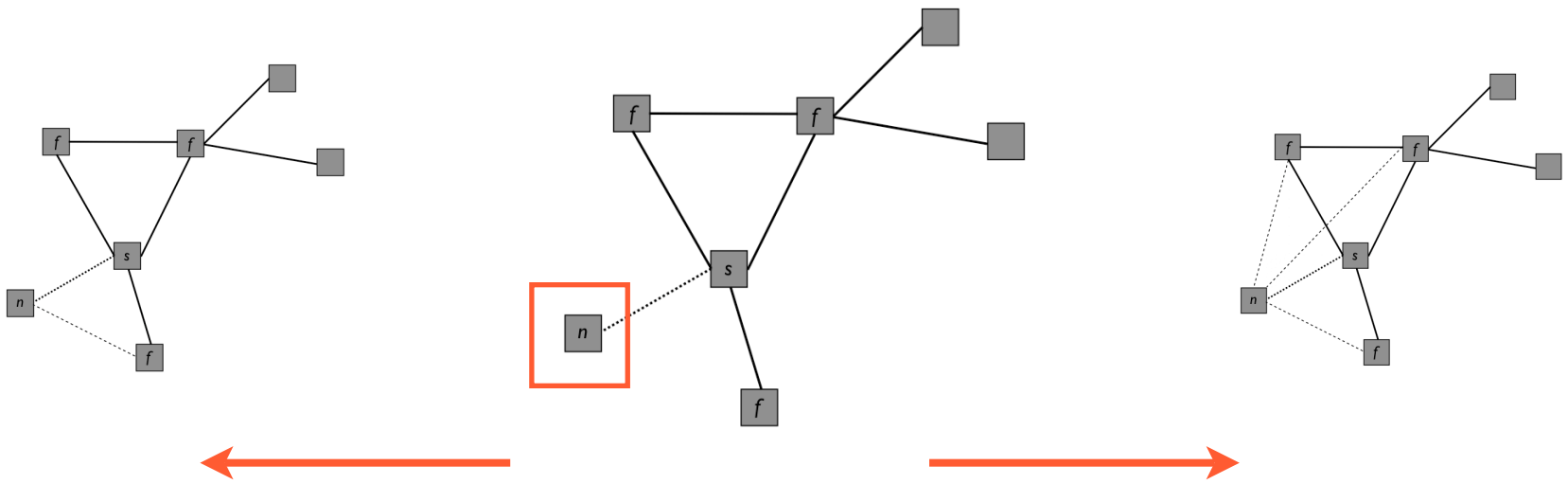




Undirected network with copying processes

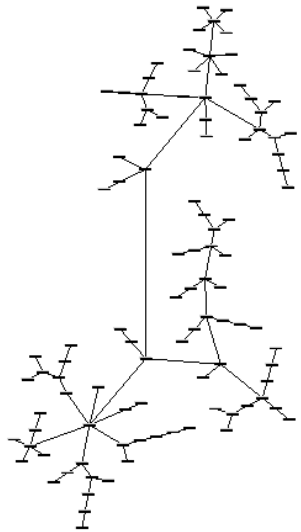
At each time step, a new node randomly connects to a target node. With probability p, it also connects to the neighbours of the target node.

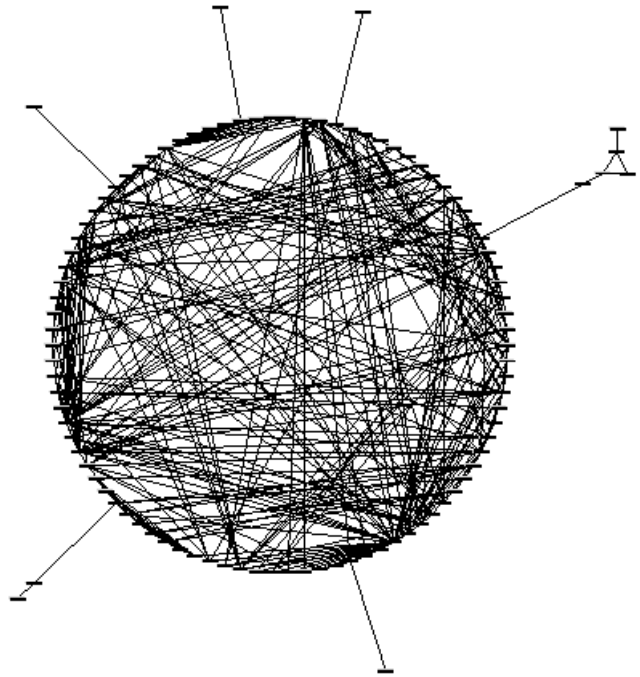
For instance, the target node already has 3 neighbours:

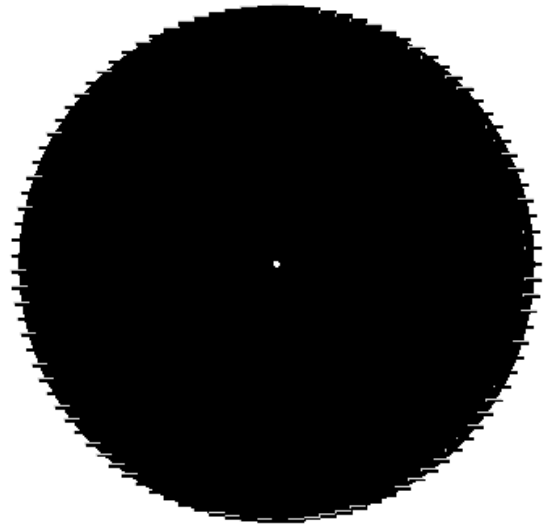


$$P = p(1-p)^2 \qquad P = p^3 \qquad \text{.....}$$

By construction, starting from *1* root node, the system grows (*N+1* nodes after N steps) and behaves like:



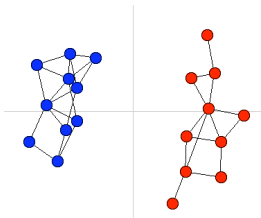


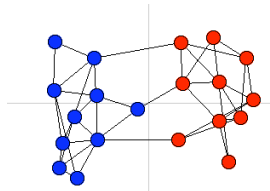


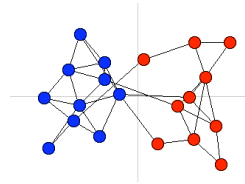
Purely random network when *p=0*

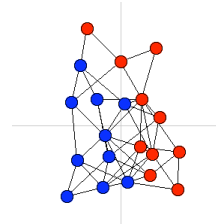
What are its properties when $p \in]0, 1[$?

Fully connected network when p=1









The system exhibits transitions for different characteristics.
For instance, the total number of links behaves like:

$$L(N) = \begin{cases} (1-2p)^{-1} N & \text{for} \quad p < 1/2; \\ N \ln N & \text{for} \quad p = 1/2; \\ A(p)\, N^{2p} & \text{for} \quad 1/2 < p \leq 1. \end{cases}$$

→ Sparse phase

→ Dense phase

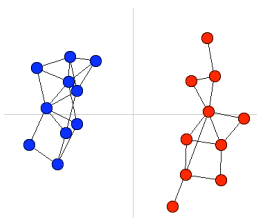
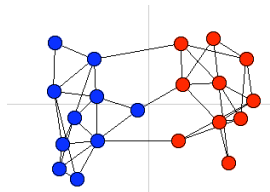
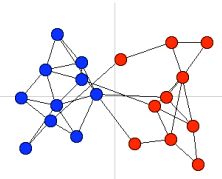
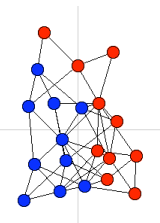
The degree distribution satisfies the non-trivial set of equations:

$$\frac{dN_k}{dN} = \frac{N_{k-1} - N_k}{N} + p\,\frac{(k-1)N_{k-1} - kN_k}{N} + m_k$$

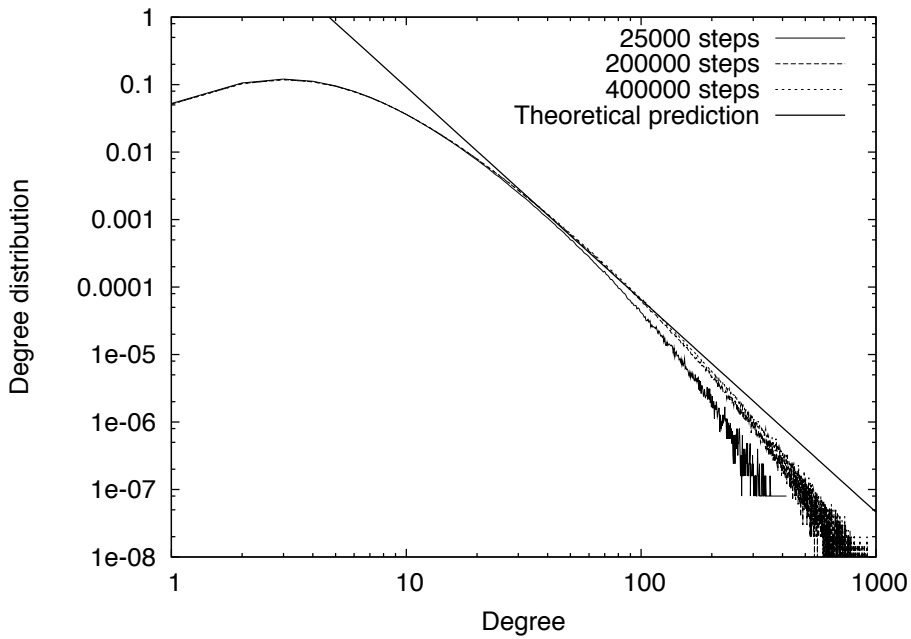
$$m_k = \sum_{s \geq k-1} n_s \binom{s}{k-1} p^{k-1} (1-p)^{s-k+1}$$

When $p < 1/2$, the asymptotic solution is stationnary:

$$n_k \sim k^{-\gamma} \quad \text{as} \quad k \gg 1$$
$$\gamma = 1 + p^{-1} - p^{\gamma - 2}$$

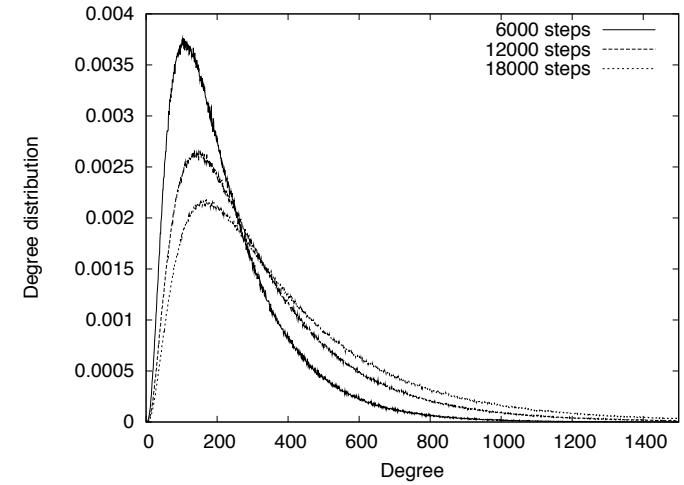





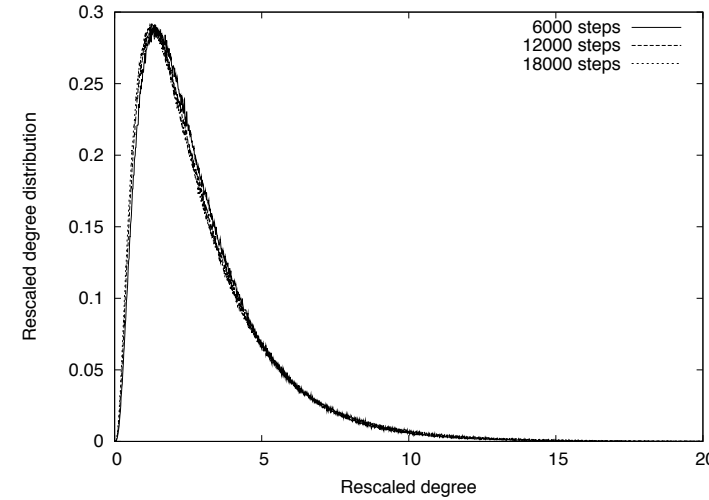
$p < 1/2$



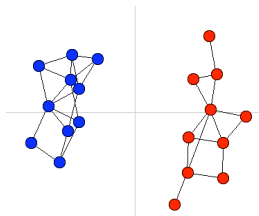
Asymptotic stationary solution

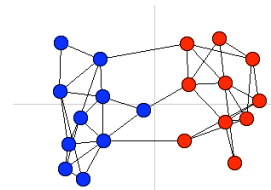
$p > 1/2$

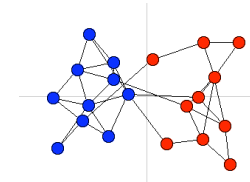


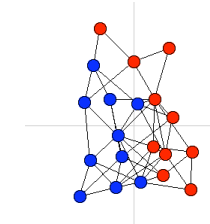


Asymptotic non-stationary self-similar solution

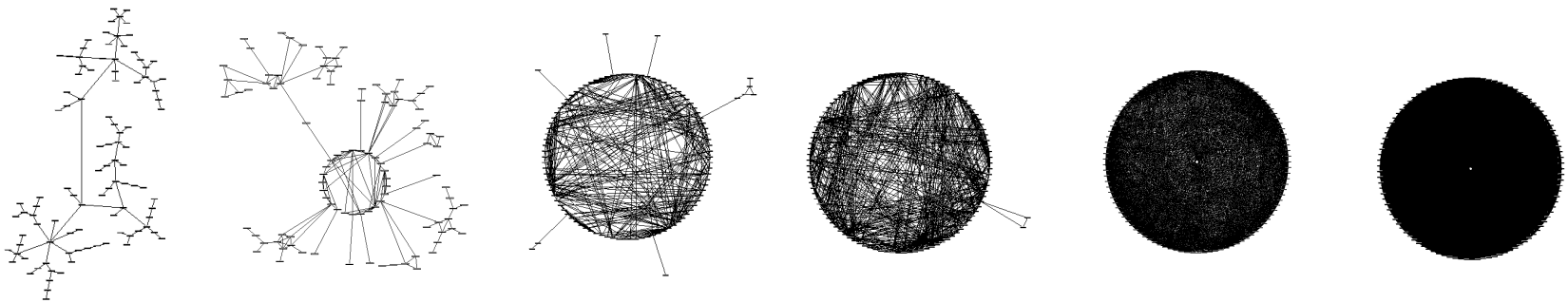




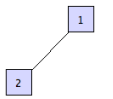




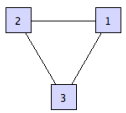
The network structure has been studied by looking at the average number of complete sub-graphs in the network



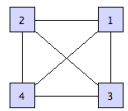
$p = 0$ $p = 0.2$ $p = 0.4$ $p = 0.6$ $p = 0.8$ $p = 1.0$



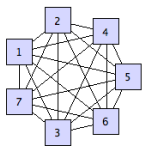
Sparse phase $L(N) = (1 - 2p)^{-1}N$ Dense phase $L(N) = A(p)N^{2p}$



$T(N) \sim L(N)$ $T(N) \sim L(N)^{\frac{3p^2}{2p}}$



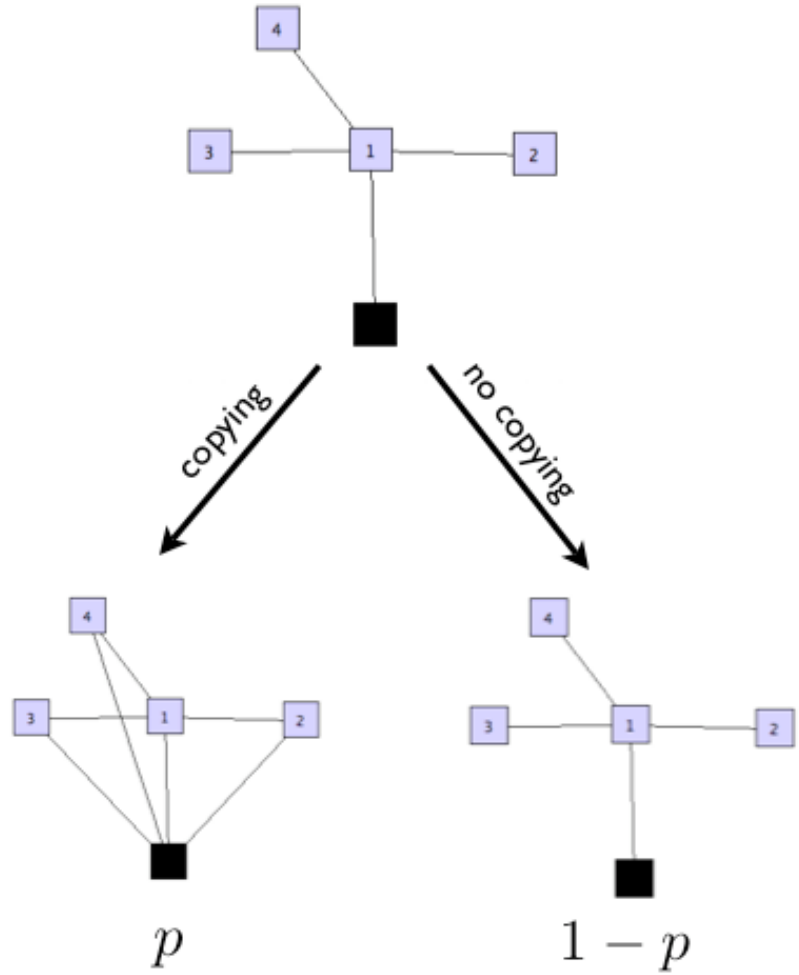
$Q(N) \sim T(N)$ $Q(N) \sim T(N)^{\frac{4p^3}{3p^2}}$

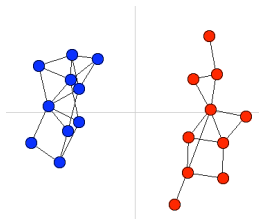


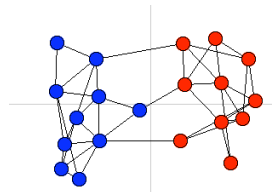
$M_m \sim M_{m-1}$ ... with transitions at $p_C = 1 - \frac{1}{m}$

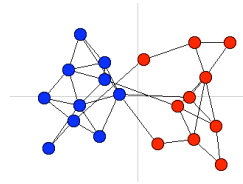
Simplified "all-or-none" model

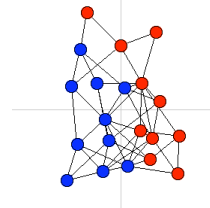
With probability p, all the neighbours of the target node are linked; otherwise, none is linked.











Distance distribution in the *"all-or-none"* version of the model

$$P(d, N+1) = \left[1 + \frac{2p}{N}\right] P(d, N) + \frac{2(1-p)}{N} P(d-1, N)$$

From which the distance moments can be found.
The average distance asymptotically behaves like:

$$d \sim 2(1-p) \ln N$$

while the fluctuations are gaussian and asymptotically vanish.

The oversimplified model reproduces all the characteristics of social models

Positive assortativity

$$r = \frac{p(9 - 14p + 11p^2)}{2 + p - 5p^2 + 2p^3}$$

Power-law degree distribution

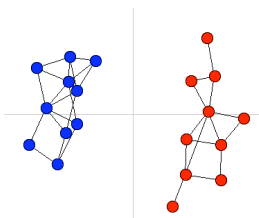
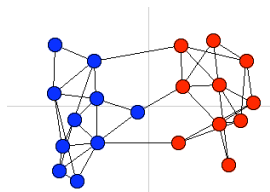
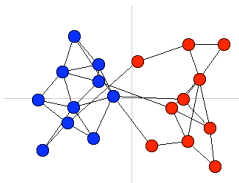
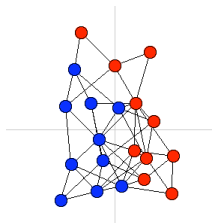
$$n_k = Y \frac{\Gamma(k + 1 + 1/p)}{\Gamma(k + 1 + 2/p)}$$

Non-vanishing clustering coefficient

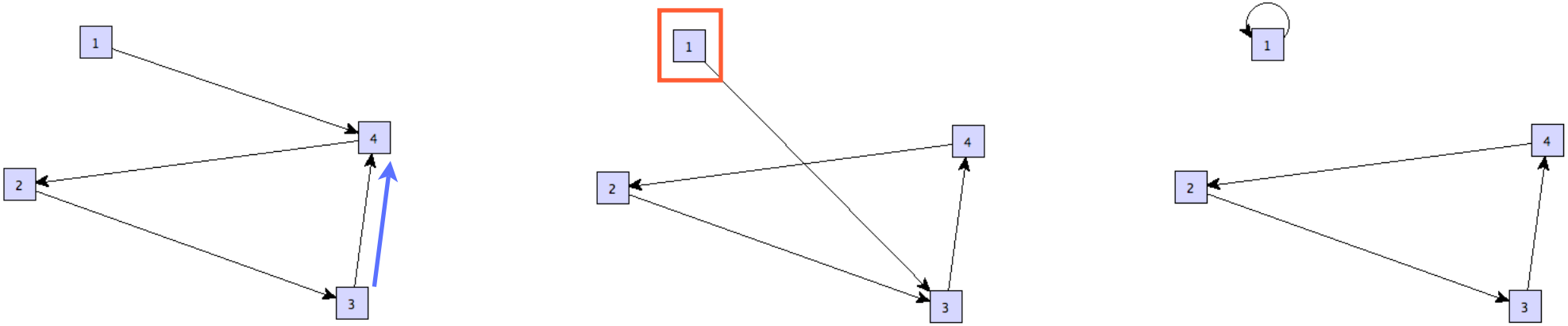
$$C = 3p/(1 + 2p)$$

Small-world behaviour

$$d \sim 2(1 - p) \ln N$$

Application to a model with redirection, with fixed number of nodes and links.



p go to the "grand-father"

$(1-p)$ connects to a node chosen randomly

small p, random structures

high p, condensation around 1 node